

AD _____

Award Number: DAMD17-01-1-0396

TITLE: Optimization of CAD System Using Adaptive Simulated
Annealing for Digital Mammography

PRINCIPAL INVESTIGATOR: Xuejun Sun, Ph.D.
Wei Qian, Ph.D.

CONTRACTING ORGANIZATION: University of South Florida
Tampa, FL 33620-7900

REPORT DATE: July 2003

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20031121 060

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE July 2003	3. REPORT TYPE AND DATES COVERED Annual Summary (1 Jul, 2002- 30 Jun, 2003)	
4. TITLE AND SUBTITLE OPTIMIZATION OF CAD SYSTEM USING ADAPTIVE SIMULATED ANNEALING FOR DIGITAL MAMMOGRAPHY			5. FUNDING NUMBERS DAMD17-01-1-0396	
6. AUTHOR(S) Xuejun Sun, Ph.D. Wei Qian, Ph.D.			8. PERFORMING ORGANIZATION REPORT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of South Florida Tampa, FL 33620-7900 E-Mail: sunxj@moffitt.usf.edu				
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited				12b. DISTRIBUTION CODE
13. ABSTRACT (Maximum 200 Words) Mammography is the most effective method to date and is becoming a high volume X-ray procedure for screening and diagnosing breast cancer. The performance of computer-aided detection and diagnosis (CAD) scheme determines its clinical effectiveness as an objective "second reader" in aiding radiologists' mammogram interpretation. Following research work of initial grant year, the major research works in the second grant year are: (1) to construct CAD system robust to FFDM and SFM, (2) to fully optimize the CAD system for its overall performance improvements in both sensitivity and specificity. The major accomplishments in the second grant year are as follows: (1) New modules have been developed, including preprocessing for normalization of mammographic images from FFDM and SFM, adaptive Fuzzy-C means algorithm for segmentation, support vector machine (SVM) technique for classification. Adaptive modules have been modified based on existing modules. (2) Adaptive CAD system has been constructed using developed and modified modules. (3) Fully optimization of CAD system by simulated annealing (SA) algorithm has been developed and performed. Key parameters affecting performance of CAD system have been selected as optimization variables. Modular and full system optimizations have been performed, respectively, on CAD system.				
14. SUBJECT TERMS Digital mammography, Computer-aided detection, optimization, Simulated annealing,				15. NUMBER OF PAGES 9
				16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

Table of Contents

Cover.....	1
SF 298.....	2
Table of Contents.....	3
Introduction.....	4
Body.....	5
Key Research Accomplishments.....	8
Reportable Outcomes.....	9
Conclusions.....	9
References.....	9
Appendices.....	none

Annual Report (Second Grant Year):

Optimization of CAD System Using Adaptive Simulated Annealing for Digital Mammography

INTRODUCTION

Mammography is the most effective method to date and is becoming a high volume X-ray procedure for screening and diagnosing breast cancer. The performance of computer-aided detection and diagnosis (CAD) scheme determines its clinical effectiveness as an objective "second reader" in aiding radiologists' mammogram interpretation. In addition, CAD should adapt to both conventional screen-film mammography (SFM) and full-field digital mammography (FFDM). The purpose of this study is to develop for the first time a new kind of fully optimized CAD system for FFDM using a global optimization algorithm to improve its performance on sensitivity and specificity in mass and MCCs detection on mammograms. Following research work of initial grant year, the major research works in the second grant year are: (1) to construct CAD system robust to FFDM and SFM, (2) to fully optimize the CAD system for its overall performance improvements in both sensitivity and specificity.

Major accomplishments of initial grant year are summarized as follows:

- (1) Databases for training and testing of CAD system performance have been constructed and corresponding truth files have been generated for FFDM and SFM respectively.
- (2) Performance of current CAD system for the detection and diagnosis of breast cancer has been retrospectively evaluated on FFDM and SFM images, respectively.
- (3) CAD modules have been being developed or modified.

Major accomplishments in the second grant year are:

- (1) New modules have been developed, including preprocessing for normalization of mammographic images from FFDM and SFM, adaptive Fuzzy-C means algorithm for segmentation, support vector machine (SVM) technique for classification. Adaptive modules have been modified based on existing modules, such as adaptive wavelet based enhancement methods.
- (2) Adaptive CAD system has been constructed using new developed and modified modules.
- (3) Fully optimization of CAD system has been developed and performed. Key parameters affecting performance of CAD system have been investigated based on FROC evaluation method, and have been selected as optimization variable. Simulated annealing (SA) algorithm has been used as optimization algorithm.

The research work of the second grant year has been accomplished successfully following the approved statement of work and progressed smoothly without technical or unexpected difficulties.

BODY

1. Development and modification of CAD modules

1.1 Preprocessing

This module is necessary and important to reduce the influence of non-breast signal, and to make CAD methods more robust and powerful and invariant to spatial resolution and gray scale conversions, since there is significant variation in gray value range and spatial resolution of mammograms from different sources due to intrinsic characteristics of different kinds of direct mammographic imaging units or screening film digitizers. Preprocessing includes histogram equalization, normalization of intensity and spatial resolution, breast area extraction, extrinsic signal elimination.

Histogram equalization algorithm has been become accepted as a standard technique for image contrast enhancement by redistribute gray level values of the pixels within an image to make the number of pixels at any gray level about the same. Its development is logical both from the point of view of optimization of information and of the local sensitivity of the human visual system. Contrast enhancement can be defined as the slope of the function mapping input intensity to output intensity. In order to improve image contrast while suppressing noise instantly, the clipped adaptive histogram equalization (AHE) has been performed [1].

As for normalization, genetic algorithm (GA) based method and nonlinear approach combined with histogram matching were previously used for geometry or intensity mapping, respectively. In this study, nonlinear intensity transformation combined with intensity stretching is conducted as described in [1]. Spatial resolution is scaled in x and y using bilinear interpolation algorithm.

1.2 Enhancement

Nonlinear filtering has proven a more robust approach than linear filtering on detail preserving, while multistage filtering was introduced in image processing in order to combine the properties of single filters and better preserve the signal and fine details of the image. In order to suppress intrinsic noise introduced in imaging process, a nonlinear tree-structured filter (TSF), a symmetric multistage filter combining the advantages of central weighted median filters (CWMF), linear and curved windows, is modified to obtain more robust characteristics for noise suppression and detail preservation [4].

1.3 Segmentation

Recognition and segmentation of mammographic lesions are extremely difficult since radiographic and morphological characteristics of lesions resemble those of tissues present in the normal breast. Lesions often do not appear as isolated densities, but are overlaid over parenchymal tissue patterns due to projection natures of mammograms. Therefore, segmentation based on pixel value alone cannot achieve accurate results. FCM, a fuzzy logic algorithm, was successfully used as segmentation approach in our previous CAD study. An adaptive FCM algorithm has been used for segmentation on mammograms. It is an iterative clustering algorithm and described in [1]. Before segmentation, DWT algorithm developed in our previous mass detection has been used to eliminate non-isotropic breast tissues to enhance isotropic regions that are most likely suspicious masses.

1.4 Classification

Nonlinear SVM algorithm has been receiving increasing research interest in recent years. It comprises a class of algorithm that represents the decision boundary in a pattern recognition problem typically in terms of a small subset of the training samples. This algorithm provides an optimally separating hyperplane with maximized margin between two classes based on the Structural Risk Minimization principle from statistical learning theory to find the optimal separating hyperplane with the lowest probability of error. Because it is a nonlinear classification problem with large dimensional feature space, SVM classification is equivalent to maximization of following dual Lagrangian function:

$$\text{Maximize: } L_D = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \phi^T(x_i) \phi^T(x_j)$$

$$\text{Subject to: } \sum_{i=1}^n \alpha_i y_i = 0 \quad \text{and} \quad 0 \leq \alpha_i \leq C$$

and sequential minimal optimization (SMO) algorithm has been implemented to speed up its optimization.

2. Full optimization of CAD system

Simulated annealing (SA) algorithm has been adopted for system optimization of CAD system. The SA algorithm is based on the principle of annealing. The annealing speed of temperature is adaptively justified to prevent the system from getting the local optimum results. The SA algorithm is a robust algorithm for finding good solutions to a wide variety of combinatorial optimization problems such as CAD system on mammograms, because it not only accepts good searching move, but also accepts bad searching move, and is independent of initial point.

The purpose of CAD is to detect and diagnose breast cancer with high sensitivity and specificity. An objective evaluation method, the normalized area, A_z , under free-response receiver operating characteristic (FROC) curve is used as the index of the system performance. It is adopted as objective function in full optimization of the CAD system.

$$F(X) = \min \frac{1}{A_z(X_i)} \quad X_i \in X$$

Three parts of SA optimization algorithm: generating function, acceptance function and cooling speed are briefly described in the following formula:

The generating function is:

$$G_c(X) = \frac{T(t)}{[T^2(t) + X^2]^{(d+1)/2}}$$

Boltzman distribution function:

$$f = Z(T) \cdot \exp\left(-\frac{F(X_i)}{KT}\right)$$

and

$$Z(T) = \frac{1}{\sum_i \exp\left(-\frac{F(X_i)}{KT}\right)}$$

The optimization procedure of full optimization of CAD system with SA algorithms has been

constructed as shown in Figure 1. Many parameters with different types in modules affect the overall performance of the CAD system. Searching step of logic, integer and continuous variables has been designed [3].

Modular and full system optimizations have been performed, respectively, on CAD system to search optimal parameter setting. Modular optimization, by optimizing variables in single module, is conducted to search parameters that significantly impact CAD performance, and to evaluate efficacy of developed optimization approach on each single module. After ranking and selecting key parameters that impact overall performance of CAD system, systematic optimization of CAD by SA algorithm has been performed.

2.1 Module optimization

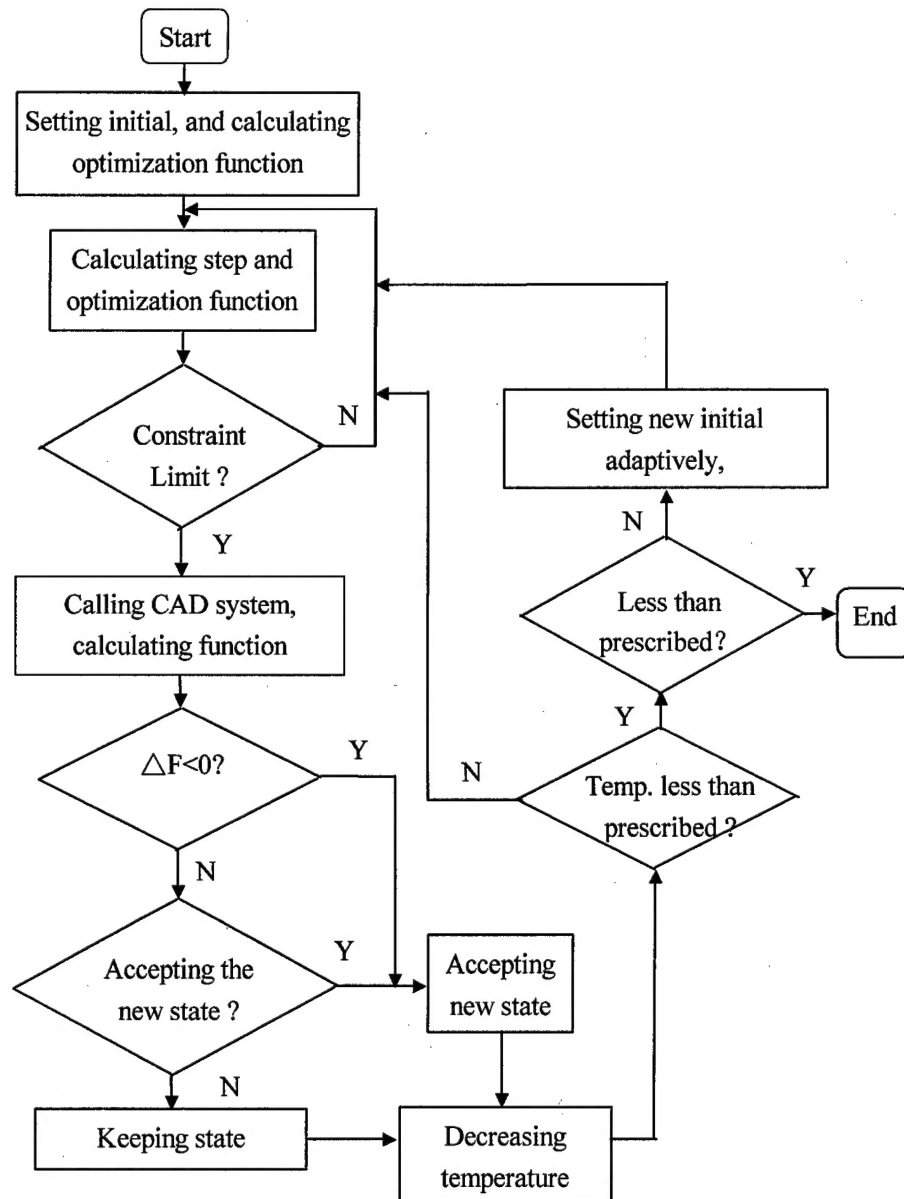


Fig. 1: Flow Chart of optimization of CAD system by SA algorithm

It has been conducted in preprocessing, enhancement, segmentation, feature extraction and selection, classification modules. Filter window size of TSF in preprocessing, sub-images selection by M-channel TSWT in enhancement, threshold of membership in FCM segmentation algorithm, and selection between adaptive clustering (AC) and FCM segmentation approaches, number of features for classification, and feature ranking and selection methods between genetic algorithm (GA) and SA algorithm, Lagrangian function in SVM algorithm, and classification algorithms between artificial neural network (ANN) and SVM, have been optimized, respectively, using module optimization approach.

Computer code for module optimization has been programmed. Obvious performance improvements have been observed through designed module optimization by proposed SA optimization algorithm. In order to speed up optimization, optimization of Boolean variables such as segmentation algorithm selection between AC and FCM, classification algorithms selection between SVM and ANN will be taken out from the system optimization procedure and processed in advance.

2.2 System optimization

Through module optimization, the following parameters have been selected as optimization variables in system optimization: window size in TSF, sub-images selection in M-channel TSWT, threshold of membership in FCM segmentation, feature number for classification, Lagrangian function in SVM classification algorithm.

This typical combinatorial system optimization on the CAD system has been conducted with constructed training database in initial grant year using designed adaptive SA algorithm. Computer code of system optimization using adaptive SA algorithm has been developed.

KEY ACCOMPLISHMENTS IN THE SECOND GRANT YEAR

- (1). Collection and generation of databases has been completed in initial grant year
- (2). Generation of Truth files for two databases has been completed in initial grant year.
- (3). Performance comparison of current CAD methods has been completed.
- (4). Development and modification of CAD modules has been completed. It makes CAD more robust and powerful, and provides foundation for full optimization of the CAD system.

Development of normalization of mammographic images makes CAD invariant to spatial resolution and gray scale conversions, being independent of characteristics of digitizer or mammographic imaging modalities and generally applicable to different kinds of mammograms. In addition, with normalization it is available to find key parameters essentially impacting overall performance of CAD system. Development of adaptive FCM segmentation and SVM classification algorithms and modification of adaptive SFM enhancement method provide more robust and adaptive approaches to CAD.

- (5). Full optimization of CAD system

Adaptive SA optimization algorithm has been developed, in which normalized area under FROC curve is adopted as objective function, constraint functions have been set to meet clinical requirements. Key parameters impacting overall performance of CAD system have been selected as optimization variables. In optimization, module and system optimization have been performed

respectively. In module optimization, parameters that affect CAD performance have been evaluated, and efficacy of SA optimization has been evaluated in each single module. Significant performance improvements have been observed through module optimization. After module optimization, full optimization of CAD system has been conducted on constructed training database using designed SA optimization approach.

REPORTABLE OUTCOME

1. Xuejun Sun, Wei Qian, and Robert A Clark, "Computer-aided diagnosis (CAD) of breast cancer on full field digital and screening film mammograms", MEDICAL IMAGING 2003, International SPIE Conference, San Diego, CA, USA, 15-20 February, 2003
2. Wei Qian, Fei Mao, Xuejun Sun, Yan Zhang, Robert A Clark, "Adaptive clustering for microcalcification detection digital mammogram", *Computerized Medical Imaging and Graphics*, No.26, pp.361-368, 2002

CONCLUSIONS

Development and modification of CAD modules makes CAD more robust and powerful. It has enriched CAD methods, and provides foundation for full optimization of CAD system. In addition, normalization of mammograms makes developed CAD methods generally applicable to both digitized film and direct full field mammograms (i.e., SFM and FFDM).

SA based full optimization has been developed for optimization of CAD system, a typical combinatorial optimization problem. Normalized area under FROC curve is used as the objective function and corresponding constraint functions to meet clinical requirements are also set. Module optimization, which is performed to rank key parameters and evaluate efficacy of proposed SA optimization approach to each single module, has been observed resulting in performance improvements. Full optimization of CAD system to selected optimization variables using adaptive SA optimization algorithm has been performed.

The research work in the second grant year has been accomplished following approved statement of work and progressed smoothly without technical or unexpected difficulties encountered.

REFERENCE:

1. Xuejun Sun, Wei Qian, and Robert A Clark, "Computer-aided diagnosis (CAD) of breast cancer on full field digital and screening film mammograms", MEDICAL IMAGING 2003, International SPIE Conference, San Diego, CA, USA, 15-20 February, 2003
2. Wei Qian, Fei Mao, Xuejun Sun, Yan Zhang, Robert A Clark, "Adaptive clustering for microcalcification detection digital mammogram", *Computerized Medical Imaging and Graphics*, No.26, pp.361-368, 2002
3. Xuejun Sun, Wei Qian, "System Oriented Optimization of CAD for Mass Detection in Digital Mammography", *Proceedings of International SPIE Conference on Medical Imaging*, San Diego, Feb. 23-28, 2002
4. W. Qian, Clarke L.P., Kallergi M., and Clark R.A., "Tree-structured nonlinear filters in digital mammography", *IEEE Trans. Med. Imag.*, 13(1): 25-36, 1994.